Data Warehousing and Data Mining

Semester: VII

Question Bank

Assignment - Section A

1. How is a data warehouse different from a database?
2. What are facts?
3. What is apex cuboid?
4. List out the components of star schema
5. What is snowflake schema?
6. List out the components of fact constellation schema?

Assignment - Section B

1. What are the factors involved while choosing data mining system?
2. Define visual data mining
3. What does audio data mining mean

Discuss the components of data warehouse

Assignment - Section C

1. What is data transformation? Why it is essential in the form of KDD? Give example
2. State why preprocessing an important issue for data warehousing and mining
3. What do data mining functionalities include?
4. Define Data Integration.
5. List the primitives that specify a data mining task
6. List the steps involved in the class comparison procedure
7. What are the uses of statistics in data mining?

Assignment - Section D

1. Explain the smoothing Techniques?
2. Explain the classifications of Data mining system.
3. .Describe challenges to data mining regarding data mining methodology and user interaction issues.
4. .Describe challenges to data mining regarding performance issues.
5. Describe issues relating to the diversity of database types.
6. Define Association Rule Mining
7. When we can say the association rules are interesting?
8. Explain Association rule in mathematical notations.
9. Define support and confidence in Association rule mining.

Important Questions - Section A

1. Write short notes on multidimensional data model?
2. Define data cube?
3. What are dimensions? Dimensions are the entities
4. Define dimension table
5. Define fact table?
6. What are lattice of cuboids?

Important Questions - Section B

1. List out the OLAP operations in multidimensional data model?
2. What is pivot operation?
3. List out the views in the design of a data warehouse?
4. What are the methods for developing large software systems?
5. Define MOLAP?
6. Define HOLAP?

Important Questions - Section C

1. Define data discretization.
2. What are the types of data pre-processing techniques? Explain in detail about them?
3. Explain Data Discretization and Concept Hierarchy Generation
4. What are the things suffering the performance of Apriori candidate generation technique.
5. Describe the method of generating frequent item sets without candidate generation.4
6. Define Iceberg query
7. Mention few approaches to mining Multilevel Association Rules

Important Questions - Section D

1. How are association rules mined from large databases?
2. Describe the different classifications of Association rule mining.
3. What are multidimensional association rules?
4. Define constraint-Based Association Mining.
5. What is Decision tree?
6. What is Attribute Selection Measure?
7. Describe Tree pruning methods.
8. Define the concept of prediction.
9. What do you mean by Cluster Analysis?
10. What are the fields in which clustering techniques are used?
11. Define CLARA and CLARANS

Short Answer Questions - Section A

1. Define data warehouse?
2. What are operational databases
3. Define OLTP?
4. Define OLAP?
5. How a database design is represented in OLTP systems
6. How a database design is represented in OLAP systems?

Short Answer Questions - Section B

1. How the operation is performed in waterfall method?
2. List out the steps of the data warehouse design process
3. Define ROLAP.

Short Answer Questions - Section C

1. Define Data mining
2. Give some alternative terms for data mining

3. What is KDD?
4. What are the steps involved in KDD process.
5. What is the use of the knowledge base?
6. What is the purpose of Data mining Technique?
7. How to generate association rules from frequent item sets?

Give few techniques to improve the efficiency of Apriori algorithm

Short Answer Questions - Section D

1. Define Genetic algorithm.
2. .Define Predictive model
3. Data mining tasks that are belongs to predictive model
4. Define descriptive model
5. Define the term summarization
6. List out the advanced database systems.
7. Define cluster analysis